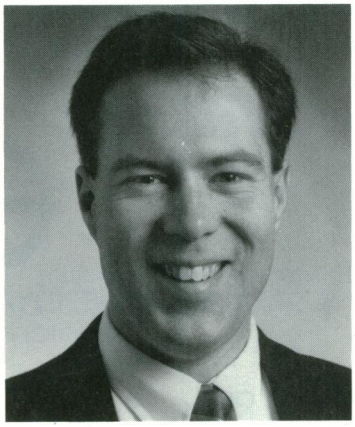
Focus

Toxicity Tests in Animals: Historical Perspectives and New Opportunities

***Editor's Note**: This article is the first of a three-part series that explores and evaluates past, current, and future uses of animals in toxicity testing. The second article will focus on alternatives to animal models, and the third will attempt to provide a balanced view of the controversies surrounding the use of animal data to predict human risks from exposure to environmental chemicals.*

Toxicology, the study of the harmful effects of substances on living systems, has been of interest since the earliest days of science. From the dregs of Socrates' last cup to the ominous silence of canaries deep between the walls of mines, toxins have been a fact of human existence. In the seemingly endless quest to define all life and ensure it against disease, hazards, and unsafe practices, animals have remained central to the search for answers from ancestry, biology, and environment. The forearms of Galen's Barbary apes answered questions about torn gladiator muscles, while his swine provided views of mammalian cardiovascular physiology which held for well over 1300 years. Centuries later, innovative breeders delighted Victorian children with new color combinations of mice, giving rise to the transgenic biology that now sheds light on some of science's most difficult questions.



Liza Green

George Gray—Animal tests are used because we want to prevent hazards.

George Gray, of the Harvard Center for Risk Analysis, cites animal models for their important role in safety evaluation. "Animal tests are used because we want to be proactive about protection from hazards; we want to prevent hazards rather than wait to find them and use humans as the experimental species."

A variety of studies define present-day toxicity testing as laboratory animal science and new biotechnologies are harnessed to address regulatory concerns about safety. The impetus for toxicity testing began in the early twentieth century as vaccines, toxins, and serums were introduced into an increasingly mobile society, one in which people moved away from direct access to fresh food, and in the direction of notoriously poor standards of food processing. Spoilage placed many people at risk of death, and although numerous scientists decried the hazards of introducing chemicals into meat and other foods, it was Upton Sinclair's *The Jungle* that helped foster the passage of the 1906 Food and Drug Act.

In addition to legislating safe practices for the preservation of food, the Food and Drug Act recognized that new vaccines carried hazards and attempted to stem the tide of quack medicines, tonics, and elixirs. Simple admonitions about the virtues of honest labeling, however, could not protect public health. In 1937, when the Massengill Company in Bristol, Tennessee, marketed one of the earliest antibiotics, sulfanilamide, with a sugar-flavored solvent of deadly diethylene glycol, more than 100 people died before the Food and Drug Administration tracked down all samples of the drug and halted its production. A consequence of this tragedy was the Food, Drug and Cosmetic Act of 1938, a landmark in drug regulation.

A number of amendments to the Food, Drug, and Cosmetic Act have been passed over the years, including the increasingly controversial Delaney Clause, which specifies that no substance that induces cancer in any animal may be incorporated into food. Debate about the clause centers around the argument that quantitative techniques are part of toxicity testing versus the belief that a single cell or molecule of a carcinogen can cause cancer. In authorizing the clause, Congressman Delaney chose the more prudent course of action. Sydney Green, director of FDA's Center for Food Safety and Applied Nutrition, criticizes the clause: "I do believe that more and more scientists are coming around to the view that we've got to do something about the Delaney Clause. Because even if there isn't a safe level per se of a carcinogen, there ought to be safe levels at which risk is almost minimal, and that's the issue. Delaney does not even allow that."

Landmark amendments were added to the Food, Drug and Cosmetic Act in 1962, after the thalidomide disaster, in which 7000 European, Canadian, and British infants born to women who took the drug were malformed. The United States had only 17 cases of such defects, largely due to the efforts of FDA Medical Officer Frances Kelsey, who kept the drug out of general distribution pending more rigorous safety testing. Despite her efforts, thalidomide was distributed to 3760 women as part of drug studies. The new Food, Drug and Cosmetic Act defined requirements for informed consent, FDA approval of clinical testing, and the important standard that new drugs demonstrate their effectiveness as well as their safety. The thalidomide disaster made the need for testing teratological effects of new drugs tragically clear.

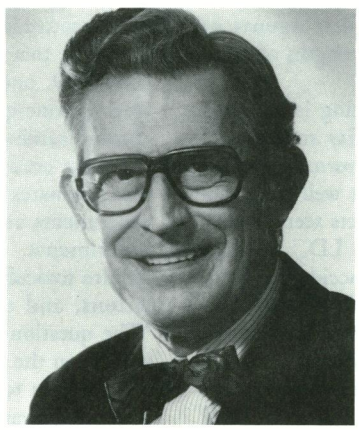
In addition to the Food, Drug and Cosmetic Act, which continues to be administered by the FDA, the 1976 Toxic Substances Act, administered by the Environmental Protection Agency, and the 1960 Federal Hazardous Substances Labeling Act, administered by the Consumer Product Safety Commission, affect toxicity testing and regulation. The federal government centralized its toxicity testing in 1978, when it established the National Toxicology Program.

To date, more than 450 chemicals have been tested by NTP, and the results have been incorporated into the Carcinogenic Potency Database (CPDB). The CPDB project began in the late 1970s at Lawrence Berkeley Laboratories, when a large body of animal data was mostly inaccessible to the scientific community. More than 4500 experiments on 1300 chemicals are now part of the CPDB, with more studies being added. "A large body of literature had accumulated in the scientific community, but there was no way to access the information easily and use it for large-scale analyses of test results. Such a capability assists in making predictions from a mouse to a rat or checking the reproducibility of bioassays in same strain, species, and sex," explains Lois Gold, director of the CPDB project. Scientists can access the CPDB before, during, and after studies to more reliably develop protocols and interpret data.

Methods and Species

Methods for exposing animals to chemicals were introduced in 1918, nearly 150 years after Joseph Priestley recorded the effects of gases on research animals. Techniques were more fully developed as scientists connected toxic exposure to diseases such as black lung in miners and scrotal skin cancer in children who routinely swept soot from eighteenth-century chimneys. Guidelines for standardized approaches to toxicity testing began to appear in the form of acute, subacute, and chronic toxicity studies in animals during the early and mid 1950s, just as pesticides and antibiotics converged on post-war American life. Exponential increases in the production of new drugs, chemicals, and environmental pollutants have compelled toxicology's race to devise and use biotechnologies, while a multitude of chemicals are screened for carcinogenic, teratogenic, and mutagenic activity, as well as toxicity to specific organic systems.

The use of *in vitro* screening procedures to test chemicals in the laboratory before they are tested in animals has alleviated some of the uncertainties associated with *in vivo* analysis. Says Louis Sibal, chief of the Office of Laboratory Animal Research at NIH, "Now we do in vitro screening tests when we're looking for new [chemotherapeutic] agents, which helps minimize the number of animals we use. We found that screening with animals didn't seem to give us very many new agents. Doing it the other way, we can pick up just as many important chemotherapeutic agents so therefore we use in vitro cultures of tumor cells."



NIEHS

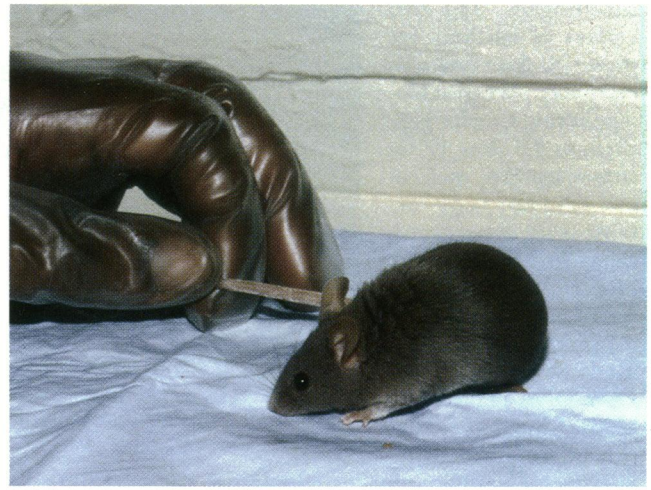
David Rall—*In vitro* tests cannot completely duplicate toxic effects.

Such *in vitro* systems have withstood rigorous tests of time and expertise. Says Sibal, "The main thing is that it's a valid way of screening, and that's something that has evolved over a long period of time. For example, the National Cancer Institute has about 22 cell cultures of human tumors growing *in vitro*, and it took years to develop all those. You can feel a little bit more secure when you screen an agent against those lines before you go on to test it in animals. Ultimately, though, we have to go into the whole animal because one cell has only a limited number of functions."

USEPA

Linda Birnbaum—Choice of animal model depends on the question being asked.

David Rall, former director of the National Institute of Environmental Health Sciences and retired assistant surgeon general, commented on the relationship between *in vitro* and *in vivo* tests: "The biochemical process leading from the administration of a chemical to the toxic effect is so complex that we really cannot duplicate it in vitro. We can duplicate bits and pieces of the process, which is why some of the mutational assays in single cells or in bacteria are very useful. In cancer, for instance, you have initiation, promotion, and so forth. You can look at any one of these in an in vitro test, but you can't put the whole thing together in a way which accurately duplicates the process in the whole organism."



ULAM, University of Michigan

No stranger to toxicity testing. Rats and mice make up 90% of the animals used in toxicity testing.

The choice of which animal to use for a particular study is vitally important. Although nearly 90% of all animals used in research are rats and mice, a number of other mammals are also used in studies, including guinea pigs, dogs, cats, rabbits, and nonhuman primates. There are many considerations to make when choosing a particular species for toxicity testing. "You choose your animal model based upon the question you are asking," explains Linda Birnbaum, director of Environmental Toxicology, EPA.

Rats and mice are often used in the early stages of toxicity testing, before it is clear what questions to ask. After some initial studies and with more data available, other species may be incorporated into studies. Because studies are done to predict the human response to a chemical based on the results from nonhuman animals, it often helps to repeat studies from one species in another. "If you get a similar kind of response, your level of confidence starts to rise that this is not only a rat response, it happens in more kinds of animals than one," says Birnbaum.

Pharmacokinetics, the study of the interactions between drugs and the body, sheds light on the discrete molecular responses to a toxic hazard such as dioxin. Ultimately, such investigations enable researchers to develop approaches that help estimate risks to humans from exposure to certain kinds of chemicals.

Pharmacokinetic studies measure the breakdown, distribution, and excretion of chemicals in the body and highlight some of the factors that regulate this process in different organ systems. Certain questions are best answered in certain species because certain animals have unique physiological variations that define their suitability for a particular investigation.

Comparing Animals with Humans

Frederica Perera, head of the Molecular Epidemiology Program at Columbia University, combined classic epidemiological techniques and bioassays to assess the effects of exposure to polycyclic aromatic hydrocarbons (PAHs) such as benzo[*a*]pyrene. Results of recent studies, published in *Nature*, were based on measurements of products of toxic exposure in people living in the Silesian region of Poland, a mining area with high concentrations of benzo[*a*]pyrene.

Bioassays are based on the fact that carcinogens, and toxins in general, enter the body and change tissues, organs, cells, and molecules. The changes, or markers, provide evidence of the effect of a particular

carcinogen. The two classes of markers are those for dose and susceptibility. Whereas biomonitoring samples populations for dose or response markers, molecular epidemiology attempts to construct a relationship between markers of dose or response and cancer risk.

In the body, PAH binds to DNA to form PAH–DNA adducts, one kind of biomarker that can be detected using immunologic, fluorescent, or radioactive labeling techniques. Other biomarkers were assessed in Perera's study, including visualizing chromosomes to look for damage and assaying oncogenes, which regulate cell growth. Perera and others are trying to derive a relationship between PAH dose or PAH–DNA concentration and the risk of cancer. "Molecular epidemiology can give us some sense of the potential risk in the human population," says Perera. "We don't want to wait to take action until we see what the health outcome is to humans."

Columbia University

Frederica Perera—Understanding how to extrapolate from animal biological markers to humans is a major goal.

Responses in animals are compared with data for humans. "Ultimately," says Perera, "we can compare the exposure to molecular dose, and dose to biological effect, and biological effect to risk relationships in both species." "But," she adds, "we want to better understand how to extrapolate from the human response in markers to those in laboratory animals for whom we know what the tumor incidence is."

Whole Animal Tests

There is a spectrum of toxicity tests using whole animals, which evaluate chemical hazards ranging from carcinogenicity to teratology and reproduction studies, as well as mutagenicity, neurotoxicity, and others. The studies can be loosely categorized as acute, subacute, subchronic, or chronic toxicity tests. One acute test, the Draize test, is used to determine whether substances that come into contact with the skin and eyes will cause irritation or injury. The eye test consists of placing drops of the test substance in one eye of as few as three test animals, generally rabbits, whose eye sensitivity is comparable to that of humans. Use of animals allows a twofold process to occur, as the eye reacts to the chemical and then begins to heal.

In the skin irritancy test, the substance in question is placed on a patch of skin that has been clipped of hair. One day later the patch is removed, and the skin is evaluated for up to three additional days. In response to concerns for animal rights, fewer Draize tests are being conducted and fewer animals are used in the tests. In addition, since its early use in the 1940s, the Draize tests have been modified and now use ophthalmic anesthetics, diluted solutions, and lower doses so as to reduce or eliminate pain or distress in test animals.

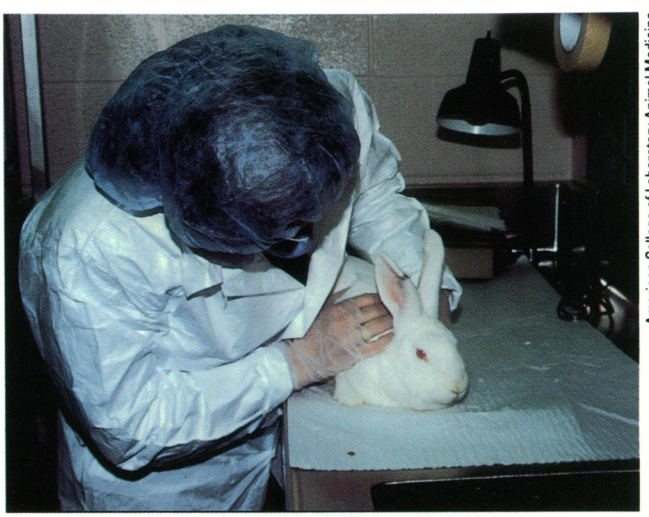
The LD_{50} or acute lethality test is generally performed in rats or rabbits and uses the dose of chemical at which one-half the test animals can be expected to die. New approaches to the classic LD_{50} are currently in practice that incorporate information from preliminary *in vitro* screening tests. The modified LD_{50}, known as the range-limit study, uses 6–10 animals instead of 80–100, as was the case before the mid-80s. The classic LD_{50} test is now generally used only to check the potency of highly toxic chemicals, such as when screening for potential chemotherapeutic agents or determining the effective strengths of pesticides.

Multiple endpoint testing is an additional part of acute lethality studies and includes observations of abnormal behavior during and after dosing, as well as autopsies that evaluate toxic effects seen in various internal organs. The LD_{50} is often viewed as analogous to an accidental exposure to toxic substances. "We need acute tests because we want to know how potent the chemical is. In other words, if people are exposed in a single accident, what sort of lethality is the chemical known to have? You can also get an indication of some of the organ systems that are affected—the brain, the kidney, and so forth," explains Rall.

Subacute and chronic studies examine the risks of longer-term exposures to chemicals and drugs and include long-term carcinogenesis studies for cancer testing. Subacute tests evaluate the effects of three different doses in two different animal species and involve administration of the substance via the same route as human exposure. Results of LD_{50} tests help determine which of the three doses given in an acute toxicity test is safe, and which dose is lethal in no more than 10% of the animals.

"The subacute or subchronic test gives a lot of information about the general toxicology: Does it affect the kidneys, bladder, brain, liver, and so on. If so, how? This all requires very careful testing, very good observation of the animals, and expensive pathology at the end of the test," says Rall. Subacute evaluations, generally 90 days long, provide the final safety tests for most consumer and household products not intended for consumption.

Chronic studies are often conducted in rats and assess the effects of long-term exposure to a substance. When humans are expected to be exposed to a chemical for the duration of a lifetime, the chronic studies extend over the animal's lifetime, which for rats is about two years. A great many chronic toxicity tests are carcinogenicity studies, in which both sexes of two species are administered the substance in question. The dose given is referred to as the maximum tolerated dose (MTD), the dose that does not shorten the life span of the animal or cause overt signs of toxicity or a weight decrease of more than 10%. NTP protocols for chronic studies generally require that two to three doses be given: the MTD, one-half the MTD, and, more recently, one-quarter the MTD. Two-year chronic carcinogenicity studies are among the most costly toxicity studies, with current estimates of $2 million for oral-dose experiments and $3 million for inhalation experiments. Moreover, chronic studies require toxicologists to make two extrapolations, and therein lie the difficulties. "The question becomes what are we finding out in the long-term studies, and what is it we want to know?" says Gold. There is increasing awareness that cancer is not nec-



American College of Laboratory Animal Medicine

Immunoglobulin factory. Rabbits are frequently used to raise antibodies in addition to being used in sensitization tests.

essarily the only health hazard that may be caused by chemical exposures. Bioassays such as those conducted by the NTP now look at other toxic effects including reproductive, developmental, immunotoxicity, and neurotoxicity.

The first extrapolation to be made in chronic studies is the more qualitative interspecies extrapolation. Because humans are rarely available, the species extrapolation has been studied by looking at the predictions between rats and mice, which are closely related. By examining the ability to predict carcinogenicity between rats and mice, and relating, as well, the target organs for incidence of tumors, Gold found that 75% of the time, if a chemical is positive for tumors for a rat, it will be positive in mice, and 50% of the time the target organ is the same in both species. Extrapolation from rats to humans is unlikely to be more accurate than that.

The second type of extrapolation relates to the fact that humans are usually exposed to low doses of carcinogens, not the high doses typified by even one-quarter the MTD. Giving low doses of cancer-causing agents to rodents would require millions of animals to detect an effect. The costs of such an approach to long-term studies are enormous, and thus protocols generally use only high doses of the chemicals being studied. Although rodent studies have provided data useful in toxicity testing, the emerging tools of molecular biology may offer some new opportunities to study the mechanisms whereby chemicals cause cancer and other adverse health effects.

New Horizons

Ray Tennant, chief of the Experimental Carcinogenesis and Mutagenesis Branch, NIEHS, has one of the first transgenic mouse lines specifically developed to investigate the molecular basis of chemical mutagenesis. "Over a decade ago Bruce Ames genetically engineered Salmonella and created a line of Salmonella that never existed before and was quite chemically sensitive. We used that tool to learn a lot about biological and chemical effects. It took a lot longer to learn how to genetically engineer the mice. But it is an equivalent process and we stand to learn even more using the genetically engineered rodents," says Tennant.

Transgenic mice are the result of advanced techniques in cell and embryo cultures, and their production involves the labor-intensive process of microinjecting DNA into the nuclei of fertilized eggs, a process which in mice may yield 20–50 animals. The value of transgenics is that investigators can change a single gene at a time, then observe the animal for changes in toxic response. This technology is "spectacularly valuable for identifying possible mechanism whereby environmental agents will cause disease," says James Huff of NIEHS's Environmental Carcinogenesis Program.

The first transgenic mice were injected with genes implicated in the development of cancer, and they showed a high incidence of certain kinds of tumors. A number of transgenic strains have been developed specifically for the purpose of studying aspects of tumor growth, including cell proliferation and differentiation.

"Cancer is a consequence of certain properties, the expression of which is highly dependent on the genotype of the animal that's exposed to a chemical. That's where transgenic biology comes in: either to insert specific genes, to regulate the expression of specific genes, or to knock out the function of certain genes. All of these provide important ways to identify critical target genes,"explains Tennant. "An enormous number of genes influence how individuals, strains or populations respond to chemicals. We have a very poor understanding of that."

The new "knockout" technology allows researchers to alter or remove a specific gene and observe the results. The tumor-suppressor gene p53 was recently deactivated in mice, with the end result that although they survived and grew, more than 70% of the mice developed tumors. Scientists think that p53 signals when DNA is damaged, causing the cell not to make more DNA until it repairs the damaged DNA. Cells change in the absence of such a signal, and the result may well be that they are transformed into tumor cells.

Not surprisingly, transgenics are difficult to create, and getting a sturdy line of mice requires extreme care. "Some of the transgenes are unhealthy and do not do well. It takes a great deal of care to come up with a line that is healthy, viable, and useful," says Tennant.

Model Care

Both traditional and transgenic animals require specialized care and attention. Many laboratory animals are special strains, with tendencies toward disease which range from specific types of tumors to increased risk of infections. "When using highly inbred mice, their resistance is down when compared to a street-caught mouse," says Birnbaum. "You can wipe out an entire colony quite quickly if you introduce some sort of common animal virus, on your shoes, for example. Even if they survive, you have compromised your data."

A variety of mechanisms ensure good animal practice in research. All Public Health Service-funded projects must comply with the Animal Welfare Act and and the Health Research Extension Act, laws that protect research animals. The policy statement, "Public Health Service Policy on Humane Care and Use of Laboratory Animals" and the well-known *Guide for the Care and Use of Laboratory Animals* provide specific methods of animal care. Institutional Animal Care and Use Committees oversee animal practices at all Public Health Service-funded facilities throughout the country. The U.S. Department of Agriculture, through its Animal and Plant Health Inspection Service, inspects all non-government facilities using animals other than rats, mice, and birds under Public Health Service funding.

The American Association for Accreditation of Laboratory Animal Care evaluates animal programs across the country. Although applying for accreditation is voluntary, more than 550 animal programs throughout the United States, Canada, and Europe are accredited through the program, ranging from industry to government to hospital facilities.

The American Association of Laboratory Animal Science coordinates aspects of its training and certification program with the American Veterinary Medicine Association. "Our students go across the board—in universities, private practice, production facilities, any place an animal lab is used," said AALAS President Richard Knauff.

As science and ethics have joined hands to develop and refine laboratory animal science, toxicologists have continued to rely on live animal models for toxicity testing. Mendelian genetics and Darwinian theories of evolution have added impetus to this reliance on animal models. From dinosaurs and giant squid to RNA-directed protein synthesis and nerve conduction, evolution has manifested a striking continuity between seemingly distinct organisms. Late twentieth-century technologies capitalize on both the sophisticated integration of whole organ systems and evolutionary continuity, with models ranging from mice and bacteria to tissue cultures and computerized simulations of whole animal physiology. Ultimately, live animal models continue to provide the knowledge necessary to protect us from toxic effects, which is toxicology's highest goal.

Mary Weideman

Mary Weideman is a freelance writer in Bethesda, Maryland.